DOCUMENT RESUME

ED 267 089 TM 860 164

AUTHOR McBride, James R.
TITLE Computerized Adaptive Testing: An Overview and an Example.
PUB DATE [85]
NOTE 11p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *Adaptive Testing; Comparative Testing; *Computer Assisted Testing; Difficulty Level; Elementary Education; *Elementary School Mathematics; Grade 6; Intermediate Grades; Mathematics Achievement; *Mathematics Tests; *Microcomputers; Testing Programs
IDENTIFIERS California; California Comprehensive Assessment System; *Computerized Adaptive Mathematics Locator; Paper and Pencil Tests; Stanford Achievement Tests

ABSTRACT

The advantages of computerized adaptive testing are discussed, and an example illustrates its use in sixth grade mathematics. These tests are administered at a computer terminal, and the test items to be administered are selected according to the difficulty level appropriate to the individual's ability. Tailoring increases the psychometric efficiency of the test by providing the same precision as a traditional test, using half as many items. Scoring is immediate and the administration procedures are easily standardized. Five essential components are the test theory, item bank, strategy, delivery system, and research. The feasibility of the Computerized Adaptive Mathematics Locator (CAML) was studied in 137 California sixth-graders. This project was part of the development of the Comprehensive Assessment System, considered as a replacement for the California Assessment Program. From a 200-item mathematics item bank, each student was given a 30-item test on the school's microcomputer, generally an Apple II. Data analysis is not yet completed. Other potential applications include military personnel aptitude testing, large-scale basic skills or minimum competency testing, diagnosis, and placement testing. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED267089

Computerized Adaptive Testing:

An Overview and an Example

James R. McBride
The Psychological Corporation

Introduction

"The test of the future!" That slogan is being used these days to call attention to computerized adaptive testing, a "new" application of computers to educational and psychological measurement.

Neither the slogan nor the application is new, however. A dozen years ago, I and a handful of fellow University of Minnesota graduate students used that slogan to attract volunteers to the first experimental trials of computerized adaptive testing. Two years ago some of those same researchers were among the participants in an AERA symposium entitled "Computerized Adaptive Testing: The Future is Now." This year, the slogan appeared in an Education Week headline, over an article describing ETS' forthcoming computerized adaptive test products.

Very soon, the first practical applications of computerized adaptive testing will be offered by test publishers for use in educational settings. By the end of the century, some advocates believe, computerized adaptive tests will completely displace paper-and-pencil tests in many testing programs.

What is computerized adaptive testing? What advantages does it offer over traditional test design and administration methods? This paper will give a brief answer to those questions. It will also assess the state of development of computerized adaptive tests, describe a prototype adaptive achievement test, and cite examples of potential educational applications. Finally, it will offer a forecast of future developments in computer administration of psychological and educational tests.

What Is Computerized Adaptive Testing?

The two features that distinguish computerized adaptive testing from traditional paper-and-pencil testing are implied in its name: computerized test administration, and adaptive test design.

TM 860 164

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. R. McBride
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The first feature is computer administration of the tests. This goes far beyond using the computer as a tool to generate printed classroom tests -- in computerized adaptive testing, the examinee takes the test at a computer terminal. Test questions are displayed on the screen, and the examinee uses a standard keyboard, or a specially designed auxiliary device, to answer them.

The second distinguishing feature is that the tests are adaptive -- individually adjusted to the ability or achievement level of the examinee. The computer program tailors the choice of test questions to reflect the examinee's performance. In effect, it chooses a more difficult question after every correct answer, and an easier question after every wrong answer. Overall, an adaptive test consists of a series of successive approximations of the examinee's ability, with each test question chosen so that its difficulty matches the current assessment of that ability level.

What Are Its Advantages?

The advantages of computer administered adaptive tests are both psychometric and practical. The psychometric benefits can be summed up in just one word: efficiency. A well-constructed adaptive test is about twice as efficient as a conventional test. Stated another way, it attains a specified level of measurement precision using about half as many test items as a conventional test would require to reach the same precision. This benefit is attributable to the adaptive feature: by tailoring the choice of questions to match the examinee's ability, the adaptive test bypasses most questions which are inappropriate in difficulty level, and which contribute little to the accurate estimation of the examinee's location on the continuum of ability.

The practical benefits of computerized adaptive testing stem both from its psychometric efficiency and from the fact that it is computer administered.

The test user can put the psychometric efficiency of adaptive tests to one of three good uses. The first alternative is to reduce the length of the test -- retaining an acceptable level of measurement precision, while cutting testing time by half on the average. The second alternative is to leave testing time unchanged, and measure with greater precision. The third alternative is to leave both testing time and measurement precision unaltered, and extend the breadth of content of the testing program.

The practical advantages of computerized test administration are almost too numerous to recite here, so I will highlight just two of them: standardization and immediacy. Clearly, administration of tests by computer has the potential to provide greater standardization of the conditions of administration than is normally the case with paper-and-pencil tests. By immediacy I am referring to the computer's inherent ability to eliminate the time-consuming and error-prone intermediate stages between test administration and the availability of the results. Immediately, and virtually without error, the computer can score an examinee's tests, convert the scores into norm- or criterion- referenced interpretations, print the results, and record them in a permanent record.

To realize the potential psychometric and practical benefits of computerized adaptive testing requires developing this technology to an advanced state. The next section is a short assessment of how far that development has progressed.

The State of the Art

What follows is my assessment of the state of the art in terms of what is needed to implement a successful computerized adaptive testing program, and how much of the needed development has been accomplished.

There are five essential components of an adaptive testing program. One of these components is a psychometric foundation -- a valid, defensible theoretical basis for administering different questions to different people, yet expressing all the results on a single scale. This foundation already exists, in the body of item response theory as advanced by Lord (e.g., 1970, 1980) and Rasch (1960; Wright & Douglas, 1975). The signal contribution of item response theory to computerized adaptive testing is that IRT provides a basis for locating both test questions and examinees on the same scale, and therefore for tailoring the difficulty of the test to the ability level of the examinee.

A second component of an adaptive test is an item bank -- a large set of test questions which measure the domain of interest and which have been calibrated using the methods of item response theory. It is desirable for the number of items in the bank to exceed substantially -- say, by a ratio of 5 or 10 to 1 -- the number of questions an individual examinee will encounter (Ree, 1977). The state of the art on this component depends on the testing program itself. In the

educational measurement arena, several large-scale programs for assessing individual achievement have already assembled large numbers of test questions which have been calibrated using the methods of item response theory. Even though adaptive versions of these achievement tests have not been developed, it would seem that the item bank resources are readily available.

A third component is a "strategy" for adaptive testing -- a set of methods for making successive approximations of the examinee's ability, and for sequentially choosing the best test question to administer at every stage of the test. Strategies of adaptive testing have been the subject of research and development for some time (e.g., Lord, 1970; Weiss, 1974; Crichton, 1981). Although there is room for still more research in this area, enough is known about some strategies to have confidence in their readiness for operational use (e.g, Wetzel & McBride, 1983).

A fourth component is a "delivery system" -- a matched set of computer hardware and software capable of presenting the test questions to the examinee, performing the computations necessary to approximate his/her location on the measurement scale, selecting the best question to present next, determining when to stop testing, and recording and reporting the results in the medium and format necessary to satisfy the requirements of the application. Experimental computer systems for adaptive testing have been available for almost two decades. Several systems capable of operational use exist today (e.g., Assessment Systems Corporation, 1984; Quan et al., 1984).

Yet a fifth component is an experience base -- a body of research, development, and empirical evidence -- which justifies confidence in the usefulness and validity of adaptive testing. This base is fairly small, but is growing rapidly along with the progress of major projects being developed by the Department of Defense and by commercial test publishers. The psychometric efficiency of computerized adaptive testing has been demonstrated several times (e.g., McBride & Martin, 1983; Moreno, Wetzel, McBride & Weiss, 1983; Urry, 1977). So has its validity as a measurement procedure (Moreno, Wetzel, McBride & Weiss, 1983) and its usefulness for predicting performance (Vicino & Hardwicke, 1984).

In summary, the state of the art is sufficiently advanced to encourage pioneering applications of computerized adaptive testing to the measurement and assessment needs of the educational community. The section that follows describes one pioneering effort.

A Prototype Adaptive Achievement Test

The California State Department of Education is developing the Comprehensive Assessment System (CAS) as a possible replacement for the California Assessment Program. The Psychological Corporation is participating in the development of CAS. As part of our CAS effort, we conducted a technology demonstration of computerized adaptive testing in the classroom.

The technology demonstration featured an experimental adaptive test of mathematics, based on the Stanford Achievement Test. We called the test the Computerized Adaptive Mathematics Locator, or CAML.

CAML was developed for the Apple II family of microcomputers, the prevailing model in use in California schools. CAML was designed for ease of use by young students. Once the computer is turned on, CAML is entirely operated by the student. All instructions are presented on the computer screen, interactively. Teacher or proctor intervention is needed only if the student has difficulty -- a rare occurrence, since the student's only task is to move a pointer on the screen. This is done by pressing one of two keys -- labelled "yes" and "no" -- in response to on-screen prompts.

CAML consisted of three 10-item adaptive subtests, so each student answered a total of 30 questions. There were over 200 mathematics achievement questions in the computer's item bank. The computer chose each question to match the test difficulty to the student's apparent level of achievement.

We administered CAML to 137 sixth-graders in three California public elementary schools in late May, using the schools' own computers. The slowest examinee finished instructions and the test in 42 minutes; the fastest one took just 17 minutes. Typical administration time was under 30 minutes.

It is too early for the statistical analyses of the CAML prototype to be complete. When all the data have been collected, we will be comparing the mathematics scores on the 30-item CAML to the 118-item Stanford mathematics scale scores. The results will be included in a report to be delivered to the Department of Education in August.

In the meantime, suffice it to note that the "test of the future" has made its first appearance in the classrooms of California schools. The next section will briefly

consider potential applications of computerized adaptive testing in the educational community at large.

Potential Applications

Three broad classes of educational measurement applications seem to be appropriate uses of computerized adaptive testing: assessment, diagnosis, and placement.

Assessment Tests

Assessment tests include aptitude or ability tests, survey achievement tests, and basic skills or competency tests.

Aptitude or ability testing is the application of computerized adaptive testing which has been most vigorously pursued. Since 1979, the Department of Defense has been preparing a computerized adaptive version of the aptitude tests it uses for selecting enlisted personnel (see Vicino and Hardwicke, 1984). In a similar vein, the Psychological Corporation is developing an adaptive version of the Differential Aptitude Tests for vocational counseling use in the schools.

Survey achievement tests could be administered in a computerized adaptive format as well. The significant advantage of that application would be the considerable savings of testing time, since achievement batteries often require four or more hours of administration time per pupil. Although this application may be appealing, it is not presently practical, since it would require a much larger number of computer terminals than is typically available in elementary or secondary schools. A related, and logistically feasible, application might be the use of adaptive "locator" tests, to identify individual candidates for out-of-level achievement testing. The CAML prototype described elsewhere in this paper is designed as a locator test.

Basic skills or competency testing is mandatory in many jurisdictions, where it is usually conducted in a large group setting. A computerized adaptive competency testing program could be conducted throughout the year, on a "walk-in" or "by-appointment" basis, or in a manner which encourages students to challenge the test when they feel ready. Again, the advantage is significant time savings per pupil, and the obstacle is the impracticality of testing large groups simultaneously.

Diagnostic Tests

An especially appealing potential application of adaptive testing is individualized diagnostic testing of academic knowledge and skills. The purpose of such tests is to identify deficiencies and prescribe remediation. This is a process which can be done much more efficiently using a computer than using the paper-and-pencil format, and which might make the computer a valuable tool to the classroom teacher.

Placement Testing

At the secondary and post-secondary levels, computerized adaptive testing may be quite useful for placement testing, by identifying strengths and weaknesses in each new student's mastery of the prerequisites of multi-level or elective courses. The Educational Testing Service is pioneering this application, with an adaptive battery of placement tests for use in community colleges (Ranbom, 1985). Similar systems may be equally useful for placement of transfer students, and students who petition for admittance to advanced placement courses.

The Future of Computerized Testing: a Forecast

Large-scale applications of computerized adaptive testing are non-existent today, for a number of reasons. Among these reasons are the cost of computers per se, and the cost of access to the fairly large mass storage capacity required to support computerized testing. In the elementary and secondary classrooms of America today, microcomputers are common, but they are usually located in computer laboratories rather than in the classrooms. Typically, these computers cost over $1,000 each, have very limited amounts of disk storage, and are not equipped to share costly common resources such as printers and large-capacity disk drives.

In the not-too distant future, I expect significant changes in the availability of computers in the classroom. These changes will include evolutionary reductions in the cost of classroom computers, revolutionary advances in the capacity of inexpensive disk storage, and increased use of local area networks to allow all the computers in a school or district to share expensive resources and to communicate with teachers' and administrators' computers.

As these changes come into being, there will be a movement towards electronic resource centers in each school or district, accessible from very inexpensive terminals or microcomputers in each classroom. These resource centers will distribute instructional courseware and testing

software for the electronic medium, at a cost comparable to that of today's textbooks and resource materials.

That evolution can be expected to take place gradually. In its later stages, I expect that every classroom will have several computer terminals or microcomputers, perhaps as many as one per pupil. These computers will not replace the teacher, but will instead serve as an electronic "workbook" which is especially useful for assessment and diagnosis of pupils, on a daily basis.

The prospect of computerized adaptive testing brings with it a challenge for educational planners to anticipate the future, to identify the most promising testing uses of computers, and to encourage the development of these tests of the future.

R E F E R E N C E S

Assessment Systems Corporation (1984). The MicroCAT testing system. St. Paul, Minnesota, author, 1984.

Crichton, L.I. (1981) Effect of error in item parameter estimates on adaptive testing (Unpublished doctoral dissertation). Minneapolis: University of Minnesota, Department of Psychology, 1981.

Lord, F.M. (1970) Some test theory for tailored testing. In W.H. Holtzman (Ed.), Computer-assisted instruction, testing, and guidance. New York: Harper and Row, 1970.

Lord, F.M. (1980) Applications of item response theory to practical testing problems. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

McBride, J.R. & Martin, J.T. (1983) Reliability and validity of adaptive ability tests in a military setting. In Weiss, D.J. (Ed.) New horizons in testing. New York: Academic Press, 1983.

Moreno, K., Wetzel, C.D., McBride, J.R. & Weiss, D.J. (1983). Relationship between corresponding Armed Services Vocational Aptitude Battery (ASVAB) and computerized adaptive testing (CAT) subtests (NPRDC TR 83-27). San Diego: Navy Personnel Research and Development Center, August 1983.

Quan, B.L., Park, T.A., Sandahl, G. & Wolfe, J.H. (1984) Microcomputer network for computerized adaptive testing (NPRDC TR 84-33). San Diego: Navy Personnel Research and Development Center, March 1984.

Ranbom, S. "E.T.S. readying computer-based 'Test of the Future'". Education Week, May 22, 1985, page 9.

Rasch, G. (1960) Probabilistic models for some intelligence and attainment tests. Copenhagen: Nielsen and Lydiche, 1960.

Ree, M.J. (1977) The effects of item calibration sample size and item pool size on adaptive testing. Applied Psychological Measurement, 1981, 5, 11-19.

Urry, V.W. (1977) Tailored testing: a successful application of latent trait theory. Journal of Educational Measurement, 1977, 14, 181-186.

Vicino F.L. & Hardwicke, S. (1984) An evaluation of the utility of computerized adaptive testing. Paper presented at the convention of the American Educational Research Association, New Orleans, Louisiana, April 1984.

Weiss, D.J. (1974) Strategies of adaptive ability measurement (Research Report 74-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, December 1974.

Wetzel, C.D. & McBride, J.R. (1983) Influence of fallible item parameters on test information during adaptive testing (NPRDC TR 83-15). San Diego: Navy Personnel Research and Development Center, April 1983

Wright, B.D. & Douglas, G.A. (1975) Best test design and self-tailored testing (Research Memorandum No. 19). Chicago: Statistical Laboratory, Department of Education, University of Chicago, June 1975.